

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

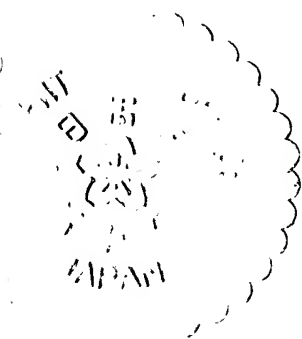
別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日            2 0 0 3 年   3 月 3 1 日  
Date of Application:

出 願 番 号            特 願 2 0 0 3 - 0 9 6 2 9 7  
Application Number:  
[ST. 10/C]:            [ J P 2 0 0 3 - 0 9 6 2 9 7 ]

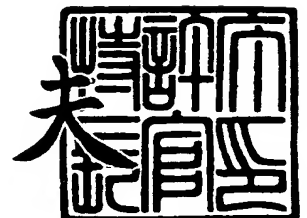
出   願   人            株 式 会 社 東 芝  
Applicant(s):



2 0 0 3 年   7 月 1 8 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



【書類名】 特許願

【整理番号】 A000205955

【提出日】 平成15年 3月31日

【あて先】 特許庁長官 殿

【国際特許分類】 H04M 11/00  
H04N 7/00

【発明の名称】 T V 電話装置、T V 電話装置のデータ送受信方法

【請求項の数】 13

【発明者】

【住所又は居所】 東京都青梅市末広町 2 丁目 9 番地 株式会社東芝青梅事業所内

【氏名】 春木 耕祐

【特許出願人】

【識別番号】 000003078

【氏名又は名称】 株式会社 東芝

【代理人】

【識別番号】 100058479

【弁理士】

【氏名又は名称】 鈴江 武彦

【電話番号】 03-3502-3181

【選任した代理人】

【識別番号】 100091351

【弁理士】

【氏名又は名称】 河野 哲

【選任した代理人】

【識別番号】 100088683

【弁理士】

【氏名又は名称】 中村 誠

## 【選任した代理人】

【識別番号】 100108855

【弁理士】

【氏名又は名称】 蔵田 昌俊

## 【選任した代理人】

【識別番号】 100084618

【弁理士】

【氏名又は名称】 村松 貞男

## 【選任した代理人】

【識別番号】 100092196

【弁理士】

【氏名又は名称】 橋本 良郎

## 【手数料の表示】

【予納台帳番号】 011567

【納付金額】 21,000円

## 【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 TV電話装置、TV電話装置のデータ送受信方法

【特許請求の範囲】

【請求項1】 ネットワークを通じて他の機器との間で映像と音声を送受信するTV電話装置において、

音声データを入力する音声入力手段と、

映像データを入力する映像入力手段と、

前記映像入力手段及び前記音声入力手段によって映像データと音声データとを入力している間にテキストデータを生成するテキストデータ生成手段と、

前記音声データと前記映像データと前記テキストデータとを合成する合成手段と、

前記合成手段によって合成されたデータを前記ネットワークを通じて送信する通信手段と

を具備したことを特徴とするTV電話装置。

【請求項2】 前記テキストデータ生成手段は、

前記音声入力手段によって入力された音声データに対して音声認識を実行する音声認識手段を有することを特徴とする請求項1記載のTV電話装置。

【請求項3】 前記テキストデータ生成手段は、

入力装置から入力されたデータをもとにテキストデータを入力するテキストデータ入力手段を有することを特徴とする請求項1記載のTV電話装置。

【請求項4】 前記合成手段は、

前記テキストデータ生成手段によって生成されたテキストデータをもとにした前記他の機器におけるテキスト表示のタイミングを調整する調整手段を有したことを特徴とする請求項1記載のTV電話装置。

【請求項5】 前記調整手段は、前記音声入力手段によって発話音声が入力された時間よりも長い間、前記テキストデータをもとにしたテキストが表示されるように調整することを特徴とする請求項4記載のTV電話装置。

【請求項6】 ネットワークを通じて他の機器との間で映像と音声を送受信するTV電話装置において、

映像データとテキストデータとが合成されたデータをネットワークを通じて受信する通信手段と、

前記通信手段によって受信したデータから前記映像データと前記テキストデータとを分離する分離手段と、

前記分離手段によって分離されたテキストデータに応じたテキストを、前記分離手段によって分離された映像データに合成する画像処理手段と、

前記画像処理手段によってテキストが合成された映像データをもとに映像を出力する映像出力手段と

を具備したことを特徴とするTV電話装置。

【請求項7】 前記画像処理手段によってテキストを前記映像データに対して合成するタイミングを調整する調整手段を具備したことを特徴とする請求項6記載のTV電話装置。

【請求項8】 ネットワークを通じて他の機器と接続されるTV電話装置において、

映像データを入力する映像入力手段と、

前記映像入力手段によって映像データを入力している間にテキストデータを入力するテキストデータ入力手段と、

前記映像データと前記テキストデータとを合成する合成手段と、

前記合成手段によって合成されたデータを前記ネットワークを通じて送信する通信手段と

を具備したことを特徴とするTV電話装置。

【請求項9】 ネットワークを通じて他の機器と接続されるTV電話装置において、

映像データとテキストデータとが合成されたデータをネットワークを通じて受信する通信手段と、

前記通信手段によって受信したデータから前記映像データと前記テキストデータとを分離する分離手段と、

前記分離手段によって分離されたテキストデータをもとに音声合成する音声合成手段と、

前記音声合成手段によって合成された音声を入力する音声入力手段と、  
前記分離手段によって分離された映像データをもとに映像を入力する映像入力手段と  
を具備したことを特徴とするTV電話装置。

【請求項10】 ネットワークを通じて他の機器との間で映像と音声を送受信するTV電話装置において、

前記ネットワークを通じて前記他の機器から前記他の機器に設けられた機能を示す情報を受信する機能情報受信手段と、

音声データを入力する音声入力手段と、

映像データを入力する映像入力手段と、

前記映像入力手段及び前記音声入力手段によって映像データと音声データとを入力している間にテキストデータを生成するテキストデータ生成手段と、

前記交換手段によって受信した前記他の機器に設けられた機能を示す情報に応じて、前記音声データと前記映像データと前記テキストデータの何れかを選択的に合成する合成手段と、

前記合成手段によって合成されたデータを前記ネットワークを通じて送信する送信手段と

を具備したことを特徴とするTV電話装置。

【請求項11】 前記ネットワークを通じて、前記他の機器に対して自装置に設けられた機能を示す情報を送信する機能情報送信手段と、

前記機能情報送信手段によって送信される機能を示す情報を設定する機能設定手段と

を具備したことを特徴とする請求項10記載のTV電話装置。

【請求項12】 ネットワークを通じてTV電話装置間で映像と音声を送受信するTV電話システムにおいて、

第1のTV電話装置では、

音声データと映像データとを入力すると共に、映像データと音声データとを入力している間にテキストデータを生成し、

前記音声データと前記映像データと前記テキストデータとを合成して、前記ネ

ットワークを通じて送信し、

第2のTV電話装置では、

前記第1のTV電話装置から送信されたデータをネットワークを通じて受信し

この受信したデータから前記映像データと前記テキストデータとを分離し、

この分離されたテキストデータに応じたテキストを、前記分離手段によって分離された映像データに合成して映像を出力することを特徴とするTV電話装置のデータ送受信方法。

【請求項13】 ネットワークを通じてTV電話装置間で映像と音声を送受信するTV電話システムにおいて、

第1のTV電話装置では、

映像データを入力すると共に、映像データを入力している間にテキストデータを入力し、

前記映像データと前記テキストデータとを合成して、前記ネットワークを通じて送信し、

第2のTV電話装置では、

前記第1のTV電話装置から送信されたデータをネットワークを通じて受信し

この受信したデータから前記映像データと前記テキストデータとを分離し、

この分離されたテキストデータをもとに音声合成して音声を出力すると共に、前記映像データをもとに映像を出力することを特徴とするTV電話装置のデータ送受信方法。

#### 【発明の詳細な説明】

##### 【0001】

#### 【発明の属する技術分野】

本発明は、TV電話装置及びTV電話装置のデータ送受信方法に関する。

##### 【0002】

#### 【従来の技術】

従来のTV電話装置は、通信回線網を介して、通信相手とするTV電話装置と

の間で映像と音声を転送する機能が設けられている。さらに、映像と音声だけでなく、付帯情報を提示できるTV電話システムも考えられている（例えば特許文献1）。このテレビ電話システムでは、電話交換システムに接続された情報提供装置において、テレビ電話機間の通信にかかる音声からキーワードを検出し、この検出されたキーワードに関連する付帯情報を記憶部から選択して、通信中のテレビ電話機に表示させる。付帯情報としては、広告情報やサービス情報で有る。広告情報とは、システムと定型している企業・商店（広告主）やその商品を宣伝するための情報である。また、サービス情報とは、ユーザに有用と考えられる様々な情報（例えば天気予報、道路マップ）のことである。

#### 【0003】

##### 【特許文献1】

特開2002-165193号公報（図1、図6、図7）

#### 【0004】

##### 【発明が解決しようとする課題】

このように従来のTV電話装置では、映像と音声を転送するだけでなく、付帯情報を表示することが考えられている。しかしながら、従来のTV電話装置では、映像と音声を利用者が使用できることを前提としているため、会話の成立に必ずしも必要ではない付帯情報を表示できるに過ぎない。例えば、聴覚が不自由な人にとっては、音声を利用することができないため、付帯情報が表示されたとしても、相手の発言内容が分からない、あるいは自分の発言を伝えにくいなど不具合があり通話を成立させるのが困難となっていた。

#### 【0005】

本発明は前記のような事情を考慮してなされたもので、映像と音声だけでなく利用者が発話した内容を表す文字列を扱うことで、聴覚が不都合な人にとっても通話を成立させることが可能なTV電話装置、TV電話装置のデータ送受信方法を提供することを目的とする。

#### 【0006】

##### 【課題を解決するための手段】

本発明は、ネットワークを通じて他の機器との間で映像と音声を送受信するT



V電話装置において、音声データを入力する音声入力手段と、映像データを入力する映像入力手段と、前記映像入力手段及び前記音声入力手段によって映像データと音声データとを入力している間にテキストデータを生成するテキストデータ生成手段と、前記音声データと前記映像データと前記テキストデータとを合成する合成手段と、前記合成手段によって合成されたデータを前記ネットワークを通じて送信する通信手段とを具備したことを特徴とする。

#### 【0007】

また本発明は、ネットワークを通じて他の機器との間で映像と音声を送受信するTV電話装置において、映像データとテキストデータとが合成されたデータをネットワークを通じて受信する通信手段と、前記通信手段によって受信したデータから前記映像データと前記テキストデータとを分離する分離手段と、前記分離手段によって分離されたテキストデータに応じたテキストを、前記分離手段によって分離された映像データに合成する画像処理手段と、前記画像処理手段によってテキストが合成された映像データをもとに映像を出力する映像出力手段とを具備したことを特徴とする。

#### 【0008】

##### 【発明の実施の形態】

以下、図面を参照して本発明の実施の形態について説明する。図1は本実施形態に係わるTV電話システム構成を示すブロック図である。TV電話システムは、複数のTV電話装置12, 14がネットワーク10を介して接続される。TV電話装置12, 14は、例えば半導体メモリ、CD-ROM、DVD、磁気ディスク等の記録媒体に記録されたプログラムを読み込み、このプログラムによって動作が制御されるコンピュータによって実現される。具体的には、パーソナルコンピュータ、PDA (personal digital assistant)、カメラ機能付き携帯電話、専用のTV電話装置などにより実現される。ネットワーク10は、例えばTCP (transmission control protocol) / IP (internet protocol) などのプロトコルを用いてデータの送受信を相互に行なうことが可能な、インターネットを含むIPネットワークである。TV電話装置12, 14には、IP (internet protocol) に従う通信機能が設けられているものとする。

## 【0009】

図2は、本実施形態におけるTV電話装置12, 14の構成を示すブロック図である。図2に示すように、本実施形態におけるTV電話装置12, 14は、音声出力部20、音声入力部22、音声処理部24、音声合成部26、音声認識部27、画像出力部28、画像入力部30、画像処理部32、テキストデータ入力部34、多重化／分離部36、通信ユニット38、機能制御部40、機能指示部42、記憶部44、及び記録再生制御部46を有している。

## 【0010】

音声出力部20は、音声処理部24から出力された音声データをもとに音声を出力するもので、スピーカなどを含む。音声入力部22は、音声を入力して音声データを音声処理部24及び音声認識部27に出力するもので、マイクなどを含む。音声出力部20と音声入力部22は、それぞれ独立してTV電話装置に設けられても良いし、ヘッドセットや受話器のように構成されても良い。

## 【0011】

音声処理部24は、多重化／分離部36からの符号化された音声データに対する復号化処理、音声入力部22から入力された音声データに対する符号化処理の他、音声合成部26によってテキストデータをもとに生成された音声データを音声出力部20から出力させる処理を実行する。

## 【0012】

音声合成部26は、多重化／分離部36によって分離されたテキストデータをもとに音声合成を行ない、合成された音声の音声データを音声処理部24に出力する。

## 【0013】

音声認識部27は、音声入力部22から入力された音声データに対して音声認識処理を実行して、例えば音声に応じたテキストデータを生成して多重化／分離部36に出力する。

## 【0014】

画像出力部28は、映像処理部32から出力された映像データをもとに映像を出力するもので、液晶ディスプレイやCRTなどの表示装置を含む。画像入力部

30は、映像を撮影して映像データを映像処理部32に出力するもので、カメラなどの撮像装置を含む。

【0015】

画像処理部32は、多重化／分離部36からの符号化された映像データに対する復号化処理、映像出力部28から入力された映像データに対する符号化処理などを実行する。

【0016】

テキストデータ入力部34は、キーボードやタブレット、マウスなどの入力装置から入力されたデータをもとにテキストデータを生成するもので、IME (Input Method Editor) などのプログラムによってテキストデータを生成する。

【0017】

多重化／分離部36は、音声処理部24（音声データ）、映像処理部32（映像データ）、音声認識部27またはテキストデータ入力部34（テキストデータ）から入力されるデータを多重化して、通信ユニット38を通じてネットワーク10に送信できるデータ形式、例えば各データをパケット化した多重ストリームデータを生成し、また通信ユニット38を介して受信されたデータから音声データ、映像データ、テキストデータを分離して、それぞれ音声処理部24、映像処理部32、音声合成部26に出力する。多重化／分離部36は、例えばMPEG (Moving Picture Experts Group) 技術を用いた多重化／分離処理を実行する。また、多重化／分離部36には、通信相手とするTV電話装置に送信したデータに応じて、送信先のTV電話装置がテキストデータをもとにテキストを表示するタイミングの調整、すなわち映像及び音声の再生とテキストの表示が同期するように多重ストリームデータの生成を調整する調整部36aが設けられている。調整部36aは、例えば、通信相手とする機器（TV電話装置）において、音声入力部22によって発話音声が入力された時間よりも長い間、テキストデータをもとにしたテキストが表示されるように調整する。

【0018】

通信ユニット38は、ネットワーク10を通じて、例えばTCP/IPによるデータの送受信を制御する。

## 【0019】

機能制御部40は、通信相手とするTV電話装置に設けられたデータ受信機能、すなわち映像データ、音声データ、テキストデータの何れに対する処理機能が設けられているかに応じて、多重化／分離部36による多重化／分離の処理を制御して、通信相手が処理可能なデータのみが送信されるようにする。機能制御部40は、通信相手のTV電話装置との間で通話が開始される前に、通信相手から処理可能な機能を示す情報が記載された機能プロフィールを通信ユニット38を通じて取得し、この機能プロフィールの記載に応じて多重化／分離部36を制御する。

## 【0020】

機能指示部42は、通信相手とするTV電話装置に対して、通話が開始される前に送信する機能プロフィール42aを機能制御部40に提供する。機能プロフィール42aに記載される情報は、TV電話装置に実装された機能に応じて固定的に決められても良いし、図示せぬ入力装置からの利用者による指示に応じて使用しない機能を任意に設定できるようにしても良い。

## 【0021】

記憶部44は、通信ユニット38を通じて多重化／分離部36に入力されたデータを記憶するもので、例えば留守番電話機能を実現するために受信データ、例えば通信相手とするTV電話装置から送信された、映像、音声、テキストの各データを記憶しておく。記録再生制御部46から再生実行が指示された場合、記録してある受信データを多重化／分離部36に提供する。

## 【0022】

記録再生制御部46は、TV電話装置を留守番電話として機能させるための制御を行なうもので、留守録設定時には多重化／分離部36を通じて受信したデータを記憶部44に記憶させ、また再生実行を指示することで記憶部44に記憶させた受信データを多重化／分離部36に提供させて、データから分離される映像(テキストを含む)、音声を出力させる。

## 【0023】

(第1実施形態)

まず、送信側のTV電話装置12において映像と音声のデータにテキストデータを付加して送信し、受信側のTV電話装置14でテキストデータをもとに文字列を表示する場合について説明する。

#### 【0024】

図3は、例えばTV電話装置12とTV電話装置14とがネットワーク10を介して接続された状態にあり、TV電話装置12からTV電話装置14に対してデータを送信する時の状況を表している。図4は、送信側のTV電話装置12の動作を説明するためのフローチャートである。図3において、受信側として示すTV電話装置14は、聴覚が不自由な人が使用しているものとする。

#### 【0025】

TV電話装置12は、映像と音声のデータにテキストデータを付加して送信する機能の実行が設定され、またTV電話装置14は、受信データ中に含まれるテキストデータをもとに文字列をキャプションとして表示する機能の実行が設定されて、それぞれ使用されるものとする。

#### 【0026】

まず、TV電話装置12は、映像入力部30によって利用者の顔などを撮影しながら、音声入力部22によって音声を入力する（ステップA1）。映像入力部30によって入力された映像データは、映像処理部32によって符号化されて多重化／分離部36に出力される。また、音声入力部22から入力された音声データは、音声処理部24により符号化されて多重化／分離部36に出力される。

#### 【0027】

一方、音声認識部27は、音声入力部22から出力される音声データを入力し、この音声データに対する音声認識処理を実行する。例えば、利用者が「こんにちは」と発話した場合には、音声認識処理によって「こんにちは」のテキストデータが生成される（ステップA2）。

#### 【0028】

図5には、映像入力部30によって入力される映像（a1）と、音声入力部22により入力される音声（発話期間）（a2）と、音声認識処理の実行期間（a3）の関係を示している。音声認識部27は、発話によって入力された音声に対

して直ちに音声認識処理を実行するので、発話が終了した時点でほぼ同時に発話内容を表すテキストデータを出力する。

#### 【0029】

多重化／分離部36は、音声処理部24から入力される音声データと、映像処理部32から入力される音声データと、音声認識部27からのテキストデータとを多重化して、多重化ストリームデータを生成するが、この際、調整部36aにより、映像及び音声に対するテキストデータの出力タイミングを調整する。

#### 【0030】

すなわち、図5(b2)に示すように、映像と音声によって確認される発話期間よりも、テキスト表示期間が長くなるように調整する。つまり、通常、音声を聞いて内容を確認するよりも、文字を読んで内容を把握する方がより多くの時間を要するためである。

#### 【0031】

多重化／分離部36は、映像及び音声に対するテキストの出力タイミングを調整した多重化ストリームデータを生成して、通信ユニット38を介して通信相手のTV電話装置14に送信する(ステップA4)。なお、多重化／分離部36は、調整された出力タイミングで、テキストデータを、映像及び音声のデータと合成して多重ストリームデータを生成しても良いし、映像及び音声に対するテキストの時間的関係を記述した関連情報を生成して、多重ストリームデータと共に送信するようにしても良い。

#### 【0032】

なお、前述した説明では、音声を入力して、音声認識処理によってテキストデータを生成しているが、テキストデータ入力部34から(キーボードなどを用いて)テキストデータを入力するようにしても良い。この場合、テキストデータ入力部34(キーボードなど)を用いたデータ入力のタイミングは、利用者が発話するタイミングと一致しない(あるいは発話しない場合もある)ので、多重化／分離部36は、テキストデータに対する出力タイミングの調整を行わないものとする。多重化／分離部36は、テキストデータ入力部34からテキストデータが入力されると、その時点で入力される映像と音声とに合成してTV電話装置14

に送信する。

#### 【0033】

次に、受信側のTV電話装置14は、通信ユニット38を介してTV電話装置12からのデータを受信すると、多重化／分離部36において、受信データから映像データ、音声データ、テキストデータを分離する。

#### 【0034】

映像処理部32は、多重化／分離部36により分離された映像データに対してテキストを合成して、映像出力部28によって出力させる。すなわち、通信相手のTV電話装置12の利用者が発声した内容、例えば図3に示す「こんにちは」の文字列が画面中にキャプションとして表示される。一方、音声処理部24は、多重化／分離部36により分離された音声データをもとに音声出力部20から音声を出力させる。

#### 【0035】

こうして、送信側のTV電話装置12において入力された音声をテキストデータに変換して送信し、さらに受信側のTV電話装置14ではそのテキストデータを画面上にキャプションとして表示するので、受信側のTV電話装置14で音声出力が有効でない場合（聴覚が不自由な人が利用しているような場合）でも、通話を成立させることができる。また、画面中に表示される文字列が、映像中で相手が発声している間よりも長い時間表示されているので、発話の内容を確実に把握することができる。

#### 【0036】

また、送信側のTV電話装置12において、音声入力を行わずにキーボードなどによるテキスト入力を行った場合であっても、受信側のTV電話装置14においてテキストを画像中でキャプションとして表示でき、送信側と受信側の何れのTV電話装置12、14とも音声入出力が有効でない場合などでも、通話を成立させることができる。

#### 【0037】

なお、前述した説明では、送信側のTV電話装置12において音声認識処理を実行し、映像及び音声と共にテキストデータを送信し、受信側のTV電話装置1

4において映像にテキストを合成して表示させているが、受信側のTV電話装置14において音声認識処理を実行するようにしても良い。この場合、送信側のTV電話装置12は、入力された映像と音声とを受信側のTV電話装置14に送信する。TV電話装置14は、多重化／分離部36によって分離された音声データを音声処理部24において復号化し、この復号化された音声データに対して、音声認識部27により音声認識処理を実行する。音声認識部27による音声認識処理により生成されたデータは、映像処理部32に出力されて映像データと合成される。映像処理部32は、映像データにテキストを合成して、映像出力部28によって表示させる。

#### 【0038】

こうして、TV電話装置12の送信側端末から送信される映像、音声に対して、受信側のTV電話装置14で音声をリアルタイムで音声認識して画面上にキャプションとして表示するので、受信側のTV電話装置14で音声出力が有効でない場合、例えば聴覚が不自由な人が使用している場合であっても、これを意識せずに、送信側の利用者はTV電話装置12を使用することができる。

#### 【0039】

次に、送信側のTV電話装置14において映像データにテキストデータを付加して送信し、受信側のTV電話装置12でテキストデータをもとに音声合成する場合について説明する。

#### 【0040】

図6は、例えばTV電話装置12とTV電話装置14とがネットワーク10を介して接続された状態にあり、TV電話装置14からTV電話装置12に対してデータを送信する時の状況を表している。図6において、送信側として示すTV電話装置14は、聴覚が不自由な人が使用している。図7は、送信側のTV電話装置14の動作を説明するためのフローチャート、図8は、受信側のTV電話装置12の動作を説明するためのフローチャートである。

#### 【0041】

TV電話装置14は、映像データにテキストデータを付加して送信する機能の実行が設定され、またTV電話装置12は、受信データ中に含まれるテキストデ



ータをもとに音声合成する機能の実行が設定されて、それぞれ使用されるものとする。

#### 【0042】

まず、TV電話装置14は、映像入力部30によって利用者の顔などを撮影しながら、音声入力部22によって音声を入力する（ステップB1）。映像入力部30によって入力された映像データは、映像処理部32によって符号化されて多重化／分離部36に出力される。また、音声入力部22から入力された音声データは、音声処理部24により符号化されて多重化／分離部36に出力される。ただし、ここでは利用者によって発声されないものとする。

#### 【0043】

一方、TV電話装置14は、テキストデータ入力部34から（キーボードなどを用いて）テキストデータが入力されると（ステップB2）、多重化／分離部36にデータデータが出力される。

#### 【0044】

多重化／分離部36は、テキストデータ入力部34からテキストデータを入力すると、その時点で音声処理部24から入力される音声と音声認識部27から入力される音声に合成してTV電話装置12に送信する（ステップB3）。なお、テキストデータが入力されなかった場合には、多重化／分離部36は、映像と音声のみをTV電話装置12に送信する（ステップB4）。

#### 【0045】

次に、受信側のTV電話装置12は、通信ユニット38を介してTV電話装置12からのデータを受信すると、多重化／分離部36において、受信データから映像データ、音声データ、テキストデータを分離する（ステップC1）。

#### 【0046】

音声合成部26は、多重化／分離部36によって分離されたテキストデータをもとに音声合成し、音声合成によって得られる音声データを音声処理部24に出力する（ステップC2）。

#### 【0047】

音声処理部24は、音声合成によって得られた音声データをもとに、音声出力

部 20 から音声を出力させる。一方、映像処理部 32 は、多重化／分離部 36 によって分離された映像データをもとに、映像出力部 28 によって映像を出力させる（ステップ C3）。

#### 【0048】

こうして、送信側の TV 電話装置 14 では音声入力を行わなくても、キーボードなどを用いたテキストデータ入力部 34 によってテキスト入力を行なうことで、受信側の TV 電話装置 12 において合成音声などによって音声として出力できるので、発話することで音声によって応答があることになり通常の会話と同じ感覚で利用することができ、また受信側にとってテキスト出力が有効でない場合（例えば、視覚が不自由な人が使用している場合）などでも、通話を成立させることができる。

#### 【0049】

##### （第 2 実施形態）

第 2 実施形態では、受信側の機器に設けられた機能が扱えるデータのみに制限して、データ送信できるようにする。

#### 【0050】

例えば、図 9 に示すように、送信側の TV 電話装置 12 は、映像、テキスト、音声の何れのデータも扱うことができる機能が設けられているが、受信側の TV 電話装置 14（例えば IP 電話装置）はテキストの表示のみしか表示できないような場合、先ず通話の前に受信側と送信側のそれぞれの機器がサポートしている機能を示す情報が記載された機能プロファイルを交換し、相互に送信できるデータ形式を認識する。

#### 【0051】

図 10 には、TV 電話装置 12 と TV 電話装置 14（IP 電話装置）との間の通話状態となるまでの手続きのシーケンスを示している。TV 電話装置 12 と TV 電話装置 14 は、ネットワーク 10 を介して接続するためのシーケンスを実行した後、相互に機能プロファイル 42a を交換する。例えば、TV 電話装置 12 は、機能制御部 40 及び通信ユニット 38 を介して、機能プロファイル 42a を TV 電話装置 14 に送信する（図 10（1））。これに対して、TV 電話装置 1

4 も同様にして機能プロファイル 4 2 a を T V 電話装置 1 2 に送信する（図 1 0（2））。

#### 【0 0 5 2】

図 1 1 には、機能プロファイル 4 2 a に記載された情報の一例を示している。図 1 1 に示す例では、映像がオフ（映像データを扱う機能が設けられていない）、音声及びテキストがオンであることを指定する情報が記載されている。

#### 【0 0 5 3】

T V 電話装置 1 2 は、T V 電話装置 1 2 から受信した機能プロファイル 4 2 a に記載された情報に応じて、機能制御部 4 0 によって多重化／分離部 3 6 が合成するデータを制限し、T V 電話装置 1 2 に送信するデータを制御する（機能設定（図 1 0（4）））。この場合、機能制御部 4 0 は、T V 電話装置 1 4 に対してテキストのみを送信するように機能設定する。

#### 【0 0 5 4】

また、T V 電話装置 1 4 も同様にして、T V 電話装置 1 2 から受信した機能プロファイル 4 2 a に記載された情報に応じて、T V 電話装置 1 2 に送信するデータを制限するための機能設定を行なう（図 1 0（3））。

#### 【0 0 5 5】

この例の場合、受信側の T V 電話装置 1 4 は、テキストを扱う機能のみをサポートしているので、送信側もそのスペックに合わせてテキストのみのデータを送信する。受信側の T V 電話装置 1 4 からは当然ながらテキストのみのデータが送信される。このように、互いにサポートしている機能が異なるような場合も通話が可能になるため、より多くの端末から通話することができるようになる。

#### 【0 0 5 6】

なお、T V 電話装置 1 2、1 4 の機能プロファイル 4 2 a に記載されている情報、すなわち図 1 1 に示す各データに対するオン／オフを、利用者からの指示に応じて自由に設定できるようにしても良い。

#### 【0 0 5 7】

すなわち、T V 電話装置 1 2 は、映像、音声、テキストの何れもサポートする機能が設けられていても、例えば映像データの送信が不要であれば、機能プロフ

ファイル 42a の映像に対する記載をオフに設定する。これにより、通信相手となる TV 電話装置 14 には、機能プロファイル 42a によって映像の機能がオフであることを通知できるので、テキストデータの受信をしなくてもすむ。

#### 【0058】

こうして、サポートしている機能を利用者自らが設定できるようにすることで、例えばネットワーク 10 の情報通信許容量が少ないような場合でも、例えばテキストと音声にデータを制限することで、柔軟的にデータ量に見合った通話を実現することができる。

#### 【0059】

なお、上述した実施形態において記載した手法は、コンピュータに実行させることのできるプログラムとして、例えば磁気ディスク（フレキシブルディスク、ハードディスク等）、光ディスク（CD-ROM、DVD 等）、半導体メモリなどの記録媒体に書き込んで各種装置に提供することができる。また、通信媒体により伝送して各種装置に提供することも可能である。本装置を実現するコンピュータは、記録媒体に記録されたプログラムを読み込み、または通信媒体を介してプログラムを受信し、このプログラムによって動作が制御されることにより、上述した処理を実行する。

#### 【0060】

また、本願発明は、前述した実施形態に限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で種々に変形することが可能である。更に、前記実施形態には種々の段階の発明が含まれており、開示される複数の構成要件における適宜な組み合わせにより種々の発明が抽出され得る。例えば、実施形態に示される全構成要件から幾つかの構成要件が削除されても効果が得られる場合には、この構成要件が削除された構成が発明として抽出され得る。

#### 【0061】

##### 【発明の効果】

以上詳述したように本発明によれば、映像と音声だけでなく利用者が発話した内容を表す文字列を扱うことで、聴覚が不自由がある人にとっても通話を成立させることが可能となる。

**【図面の簡単な説明】**

【図 1】 本実施形態に係わる TV 電話システム構成を示すブロック図。

【図 2】 本実施形態における TV 電話装置 12, 14 の構成を示すブロック図。

【図 3】 TV 電話装置 12 から TV 電話装置 14 に対してデータを送信する時の状況を表す図。

【図 4】 送信側の TV 電話装置 12 の動作を説明するためのフローチャート。

【図 5】 映像 (a1) と音声 (発話期間) (a2) と音声認識処理の実行期間 (a3) の関係を示す図。

【図 6】 TV 電話装置 14 から TV 電話装置 12 に対してデータを送信する時の状況を表す図。

【図 7】 送信側の TV 電話装置 14 の動作を説明するためのフローチャート。

【図 8】 受信側の TV 電話装置 12 の動作を説明するためのフローチャート。

【図 9】 送信側の TV 電話装置 12 と受信側の TV 電話装置 14 が扱うことができる機能を示す図。

【図 10】 TV 電話装置 12 と TV 電話装置 14 (IP 電話装置) との間の通話状態となるまでの手続きのシーケンスを示す図。

【図 11】 機能プロファイル 42a に記載された情報の一例を示す図。

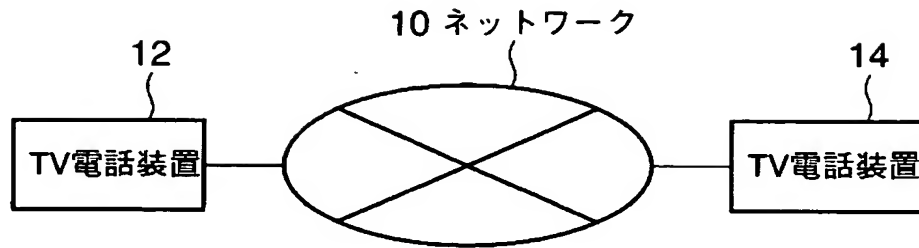
**【符号の説明】**

10…ネットワーク 10、12, 14…TV 電話装置 14、20…音声出力部 20、22…音声入力部 22、24…音声処理部 24、26…音声合成部 26、27…音声認識部 27、28…映像出力部 28、30…映像入力部 30、32…映像処理部 32、34…テキストデータ入力部 34、36…多重化／分離部 36、36a…調整部 36a、38…通信ユニット 38、40…機能制御部 40、42…機能指示部 42、42a…機能プロファイル 42a、44…記憶部 44、46…記録再生制御部 46。

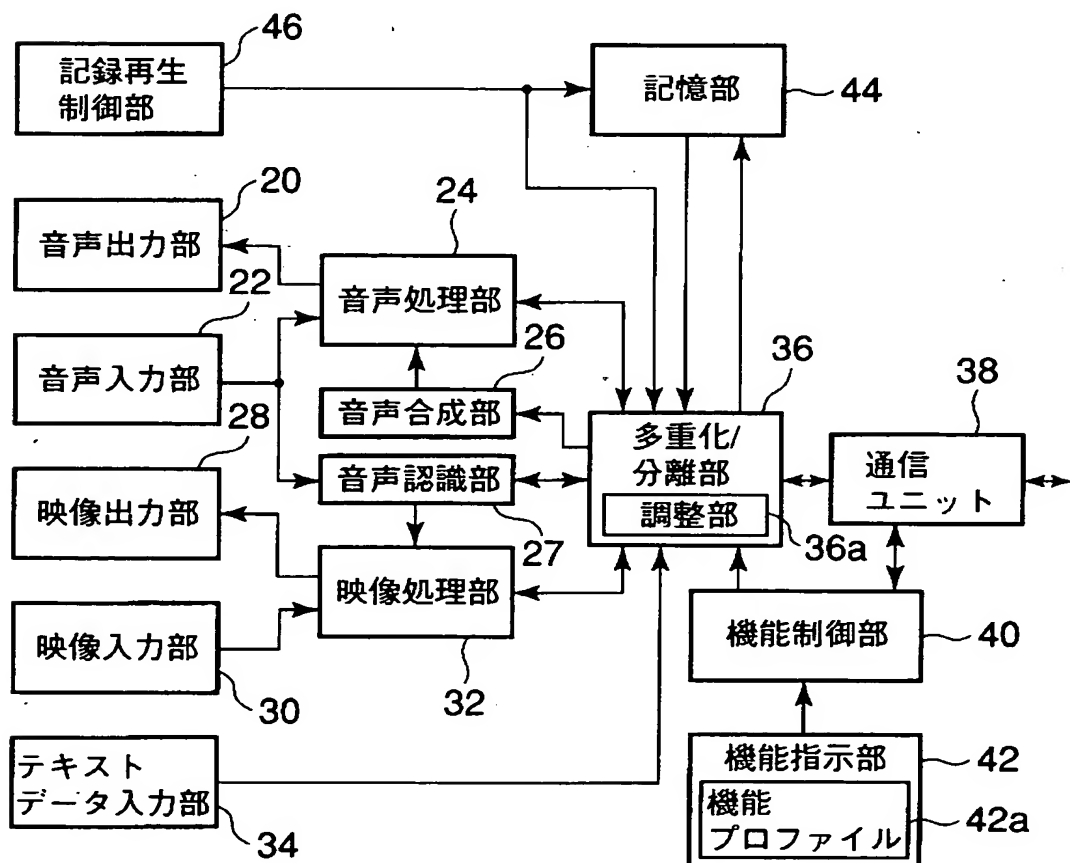
【書類名】

図面

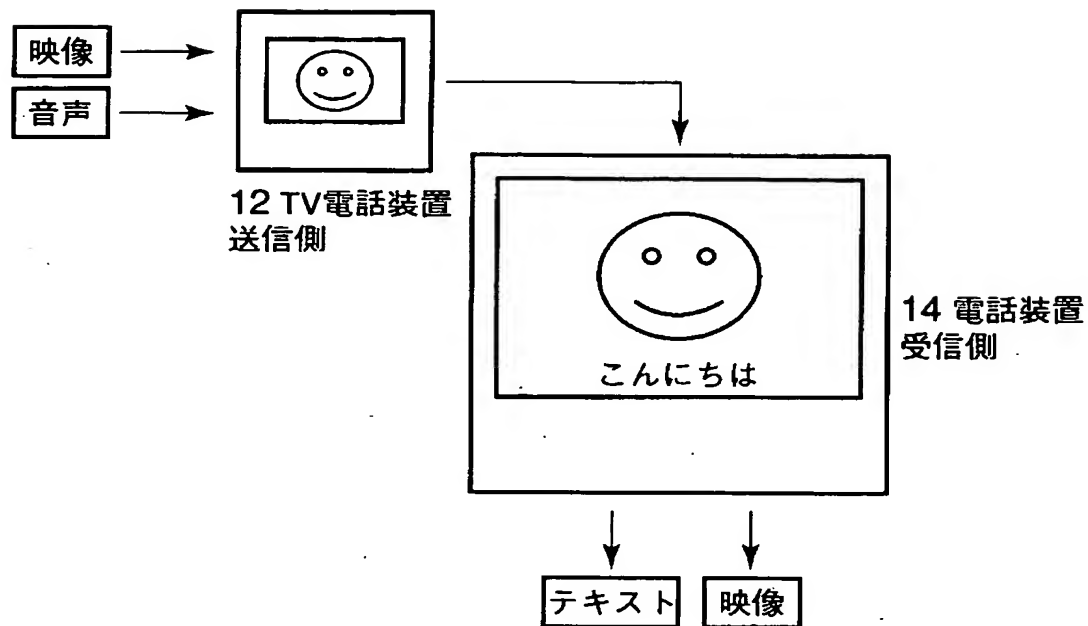
【図 1】



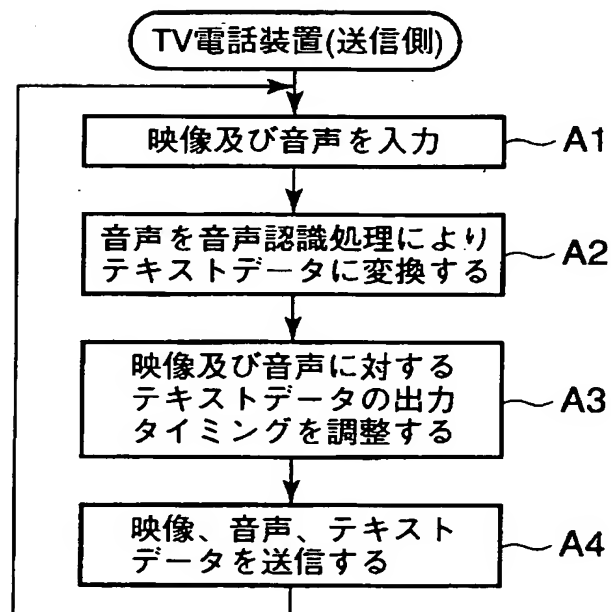
【図 2】



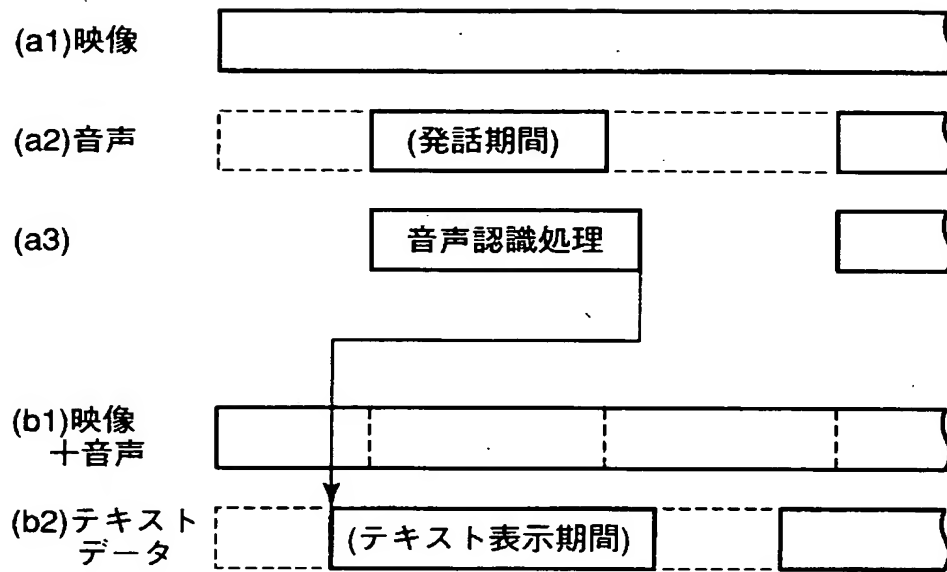
【図 3】



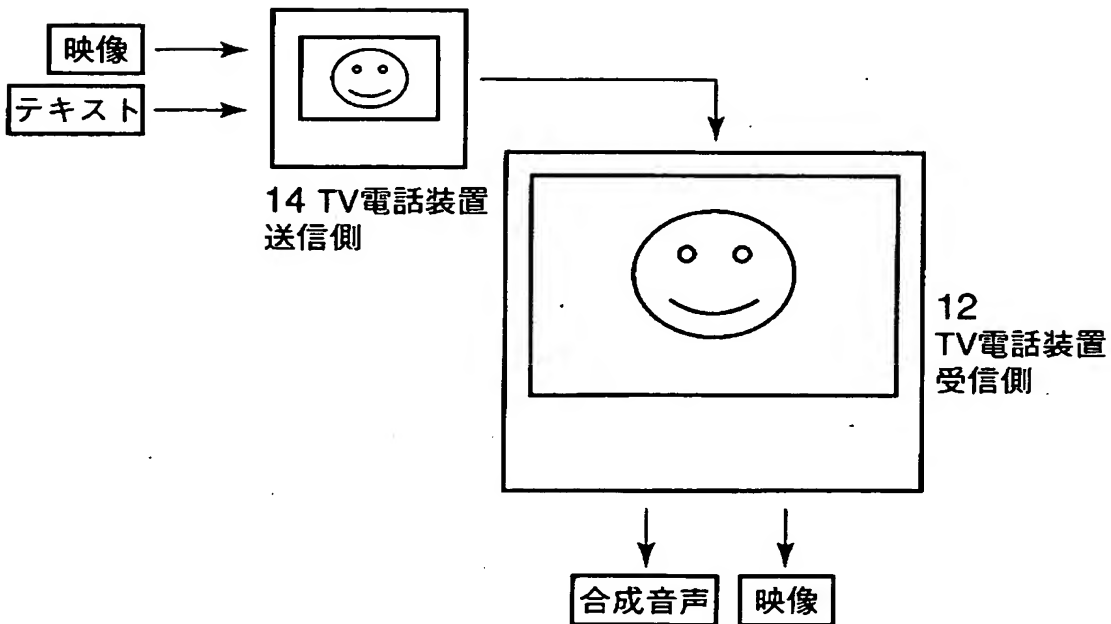
【図 4】



【図 5】

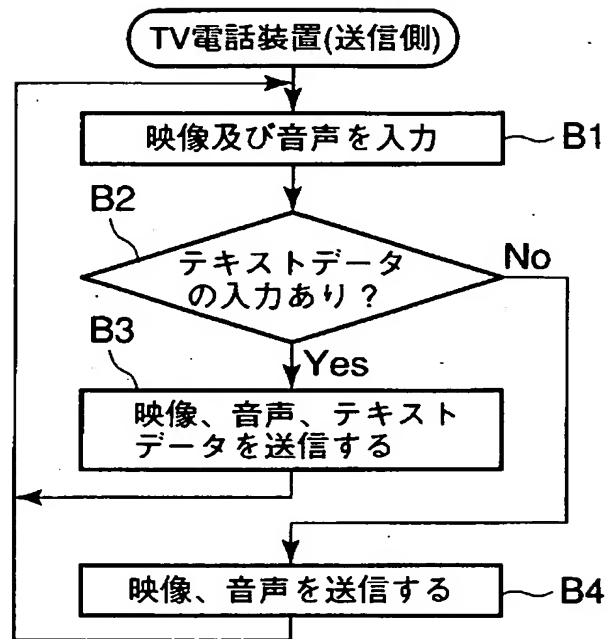


【図 6】

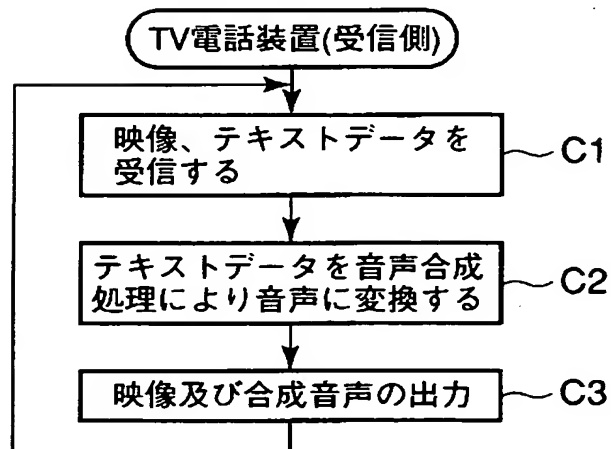




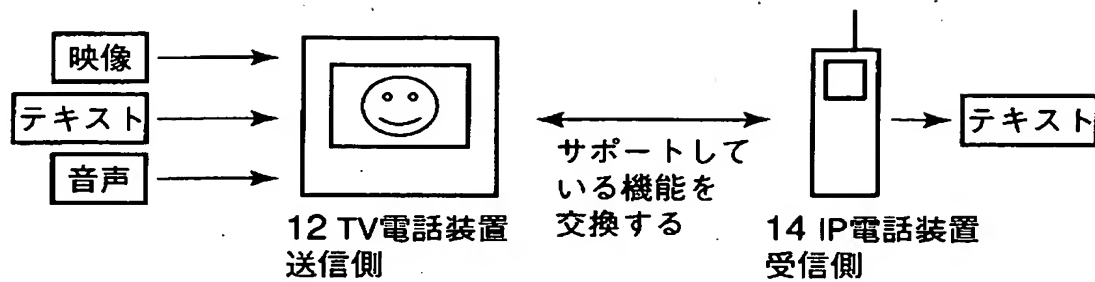
【図 7】



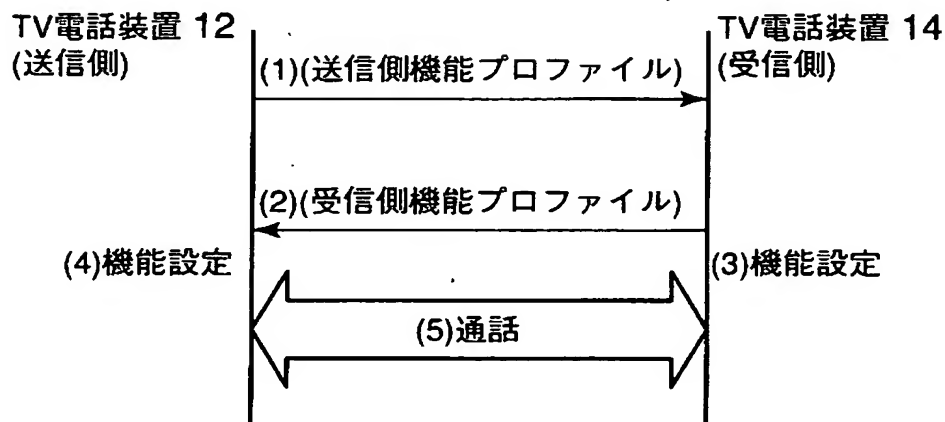
【図 8】



【図 9】



【図 10】



【図 11】

機能プロファイル	
映像	オフ
音声	オン
テキスト	オン

【書類名】 要約書

【要約】

【課題】 映像と音声だけでなく利用者が発話した内容を表す文字列を扱うことで、聴覚が不自由な人にとっても通話を成立させる。

【解決手段】 ネットワークを通じて他の機器との間で映像と音声を送受信するTV電話装置である。TV電話装置は、音声データを入力する音声入力部22と、映像データを入力する映像入力部30と、入力された音声データに対して音声認識処理を実行してテキストデータを生成する音声認識部27と、音声データと映像データとテキストデータとを多重化する多重化／分離部36と、多重化されたデータをネットワークを通じて送信する通信ユニット38を有する。

【選択図】 図1

特願 2003-096297

出願人履歷情報

識別番号

[000003078]

1. 変更年月日

2001年 7月 2日

[変更理由]

住所変更

住 所

東京都港区芝浦一丁目1番1号

氏 名

株式会社東芝

2. 変更年月日

2003年 5月 9日

[変更理由]

名称変更

住所変更

住 所

東京都港区芝浦一丁目1番1号

氏 名

株式会社東芝